Technical Report 466

LEVEL II

13

EFFECTS OF HUMAN SELF-ASSESSMENT RESPONDING ON LEARNING

AD A109455

Darwin P. Hunt
New Mexico State University

U. S. Army
Research Institute for the Behavioral and Social Sciences

August 1980

82 01 11 139

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

FRANKLIN A. HART
Colonel, US Army
Commander

JOSEPH ZEIDNER
Technical Director

Research performed under contract to
the Department of the Army

Psychology Department
New Mexico State University

NOTICES

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report 466	2. GOVT ACCESSION NO. AD-A109455	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) EFFECTS OF HUMAN SELF-ASSESSMENT RESPONDING ON LEARNING		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report
		6. PERFORMING ORG. REPORT NUMBER --
7. AUTHOR(s) Darwin P. Hunt		8. CONTRACT OR GRANT NUMBER(s) DAH19-76-G-0001
9. PERFORMING ORGANIZATION NAME AND ADDRESS Psychology Department New Mexico State University Las Cruces, NM 88003		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS US Army Research Institute for the Behavioral & Social Sciences 5001 Eisenhower Ave., Alexandria, VA 22333		12. REPORT DATE August 1980
		13. NUMBER OF PAGES 64 48
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) --		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE --

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

Training
Self-Assessment Responding
Confidence Rating
Subjective Report
Human Performance
Information Processing
Practice

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Training
Self-Assessment Responding
Confidence Rating
Subjective Report
Human Performance
Information Processing
Practice

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A conceptual framework of a process by which persons assess and express levels of sureness in the correctness of responses which they anticipate making - or have already made but not yet received knowledge of results - is proposed. It is hypothesized that the rate at which a person's behavior is modified by knowledge of results is affected by the covert and overt sureness associated with the execution of responses which are being learned. Data are presented which show that acquisition in a paired-associates learning task may be enhanced by the concomitant performance of a self-assessment (SA) task.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

406684

Acquisition is more rapid and the self-assessments regarding levels of sureness are more accurate if the response to-be-learned is executed before the SA response rather than after it. Learning involved a decrease in the proportion of unsure-wrong responses and an increase in the sure-correct responses, with little change in the proportion of sure-wrong and unsure-correct responses. Based upon an analysis of the relation between the sureness/correctness of the responses and the speed with which responses are executed, it is suggested that the process and/or factors involved in determining the correctness of a response may be different from those involved in determining its sureness.

Technical Report 466

EFFECTS OF HUMAN SELF-ASSESSMENT RESPONDING ON LEARNING

Darwin P. Hunt
New Mexico State University

Milton H. Maier, Contracting Officer's Representative

Submitted by:
Robert M. Sasmor
Director, Basic Research

Approved by:
Joseph Zeidner
Technical Director

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

August 1980

Army Project Number
2Q161102B74F

Basic Research in
Training

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The Training Technical Area of the Army Research Institute for the Behavioral and Social Sciences (ARI) conducts basic research in support of the systems engineering approach to training. The major focus of this research is to develop fundamental data and technology for improving individual job performance. This report is one of a series on specific topics in the area of skill acquisition and retention. It discusses the effects of a learner's self-assessment and indication of confidence in an answer on how effectively the lesson is learned. Research was conducted at New Mexico State University under grant DAHC19-76-G-0001 and was monitored by Milton H. Maier as part of Army Project 2Q161102B74F. J. V. Bradley, N. S. Urquhart, and G. M. Southward of New Mexico State University provided statistical assistance. The working environment at the Georgia Institute of Technology, where the author was a visiting professor for 1977-78, encouraged the research and beneficially affected preparation of this report.

JOSEPH ZEIDNER
Technical Director

EFFECTS OF HUMAN SELF-ASSESSMENT RESPONDING ON LEARNING

BRIEF

Requirement:

To determine the effects of self-assessment (SA) responding on the rate of learning. SA responding requires the learner to indicate a level of sureness in the correctness of each answer given.

Procedure:

Nine different groups of 20 students each were given the primary task of learning the names of eight different pairs of pliers. Line drawings of the pliers were projected on a screen one at a time. Students answered by pressing a labelled button. Pictures were presented in different sequences until the students could name all eight pliers correctly on two consecutive trials.

In six experimental groups, students indicated how sure they were about the correctness of their answer by pressing one (of two, four, or eight) SA-response buttons after an answer had been made. Three experimental groups made their SA-response before pressing an answer button and three did so after pressing an answer button. The number of trials the experimental groups needed to learn the material was compared with the number of trials needed by a control group, which performed only the primary task of learning the plier's names, or by two other groups who pressed a single available button labelled "Record" either before or after answering.

Findings:

Students in the SA group who (a) made their SA response after each answer and (b) used eight SA-response buttons required an average of 25.3% fewer trials to learn the material than did those in the control group who performed only the primary learning task (20.5 vs 15.3 trials). Making the SA response after each answer benefited learning more than making it before each answer. SA responding seemed especially helpful to the slower learners.

The speed of correct responses (but not wrong responses) was affected by the associated sureness. Sure-and-correct responses were made an average of about one second faster than unsure-but-correct responses. Wrong answers took an average of about five seconds, regardless of sureness. Sure-but-wrong answers took about 1.6 seconds longer than sure-but-correct answers.

Students in the two "Record" groups also learned faster than the control group. Apparently, the benefits of SA responding are not solely due to the cognitive component of self-assessment but may also involve the motor component. A detailed conceptual model of the human self-assessment process is proposed which relates SA responding to learning.

Utilization of Findings:

The validity and reliability with which persons can assess their own knowledge of task performance have an important effect on human performance and training. A person's decisions as well as the latency, speed, vigor, and smoothness of responses may be directly related to this self-assessment process.

The findings show that it is possible to expedite learning in at least some identification tasks by the appropriate use of SA responding during training. It should be relatively easy to apply these findings to some operational training situations to evaluate the practical merits of SA responding. However, additional research is needed to (a) verify the findings, (b) identify more precisely and with more confidence the factors in SA responding which expedite learning and the ways in which training and SA responding interact, and (c) define the domain of tasks whose training can and cannot benefit from SA responding.

EFFECTS OF HUMAN SELF-ASSESSMENT RESPONDING ON LEARNING

CONTENTS

List of Tables

List of Figures

EFFECTS OF HUMAN SELF-ASSESSMENT RESPONDING ON LEARNING

INTRODUCTION

> The <u>proud</u> man . . . is an extreme in respect of the greatness of his claims, but a mean in respect of the rightness of them; for he claims what is in accordance with his merits, while others go to excess or fall short.
>
> . . . he who thinks himself worthy of great things, being unworthy of them is <u>vain</u>.
>
> Aristotle, 4th century B.C.
> (translator W. D. Ross in Auden, 1970)

It is widely accepted that human performance is affected by the knowledge which an individual has stored in memory, by the rapidity and accuracy with which such knowledge may be retrieved and processed, and by whether the responses required to translate a decision into action can be appropriately selected and executed. The main point of this paper is that the performance of an individual also importantly depends upon the validity and reliability with which the person can assess whether items of knowledge and responses which are relevant to the performance of the task are stored in his/her own memory, are retrievable from it and are executible.

If an individual is given a choice as to whether to engage in some task or activity, such as driving an automobile, the decision of the person as well as the manner in which the task is executed depends not only upon whether the person possesses the knowledge and capacities necessary to perform the activity but also upon the person's self-assessment of whether (and the extent to which) the knowledge and capacities are possessed by him.

Furthermore, this self-assessment (SA) process may interact with the processes by which knowledge is acquired and retained in memory. That is, learning may be influenced by the manner in which the SA process is involved during the period of time when knowledge and responses are being acquired and retained.

The processes by which such self-assessments are accomplished by an individual and some ways in which learning may interact with the SA process is the topic of this paper. These processes, the components and their interactions are of both theoretical and practical importance. The effects of the SA process should be reflected in the spatio-temporal characteristics, e.g., latency, vigor, and smoothness, of motor and verbal responses.

First a conceptual framework (which, for brevity, is called a model) within which to consider the SA process and learning is presented. Then some data are presented and discussed concerning (a) the effects of performing a self-assessment task on the rate of learning in a paired-associates learning task, (b) some changes which occur in the self-assessment responses with practice, (c) the order in which the self-assessment responses and the responses to-be-learned are covertly or

internally selected, and (4) the accuracy of the self-assessment responses in the paired-associates learning task.

A MODEL OF A SELF-ASSESSMENT PROCESS

The model which is diagrammed in Figure 1 is intended to provide a framework within which to consider the details of a self-assessment process. It is not intended necessarily to portray the underlying physiological [illegible] involved. The proposed model involves an item-by-item iterative [illegible] is considered to be only one manner in which some level of sure-[illegible] the correctness of some anticipated or executed responses) may be [illegible]oduced by an individual. An alternative is that the degree of sureness may, in some instances, be based upon a general information memory (Nuttin & Greenwald, 1968) rather than a retrieval and testing of specific items, responses, etc. Another alternative which might be operative under some circumstances--or perhaps may be the first of a two-stage SA process--would involve the person having direct access to some items (Kolers & Palef, 1976).

The model presented here borrows specific concepts and approaches from Kelley (1968); Miller, Galanter, and Pribram (1960); Adams (1971); and Attneave (1974). The components of the proposed model of the self-assessment process will be considered separately, but it will be useful to summarize the manner in which the model of the total process is envisioned to function. Generally the capital letters indicate events which are observable (such as overt responses of the person) and the small letters indicate internal, implicit or covert responses, events, or states.

1. Based upon the individual's perception ($\underline{s}$) of the Situation (S) and upon the Goal of the individual, Internal Models ($\underline{s} \rightarrow \underline{m_i} \rightarrow \underline{c_m}$) of the real-world and specific responses ($\underline{m_i}$) are retrieved from memory.

2. The consequences ($\underline{c_m}$) predicted covertly as a result of inserting the selected $\underline{m_i}$ into the retrieved Internal Model are compared cognitively with the consequences desired as implied by the Goal.

3. The closer the agreement between the desired consequences and the predicted consequences then the higher the sureness, $\underline{k}$, of the individual in the correctness of the $\underline{m_i}$ and the Internal Model. A close match between the predicted consequences and the desired consequences produces a high level of sureness that the knowledge necessary to perform some act correctly is stored--and that the act if performed under the perceived situation will result in certain desired consequences.

4. This sureness is then tested against a Criterion-$\underline{k}$. If the criterion is met or exceeded--and if the individual determines that the response can be executed successfully[1]--then $\underline{m_i}$ is executed ($\underline{M}$). Otherwise the $\underline{m_i}$ (and/or the Internal Model) is rejected and a new

---

[1]The individual's estimation of whether the $\underline{m_i}$ response can be executed successfully (Bandura, 1977) is viewed as separate from the self-assessment of the correctness or appropriateness of a response.

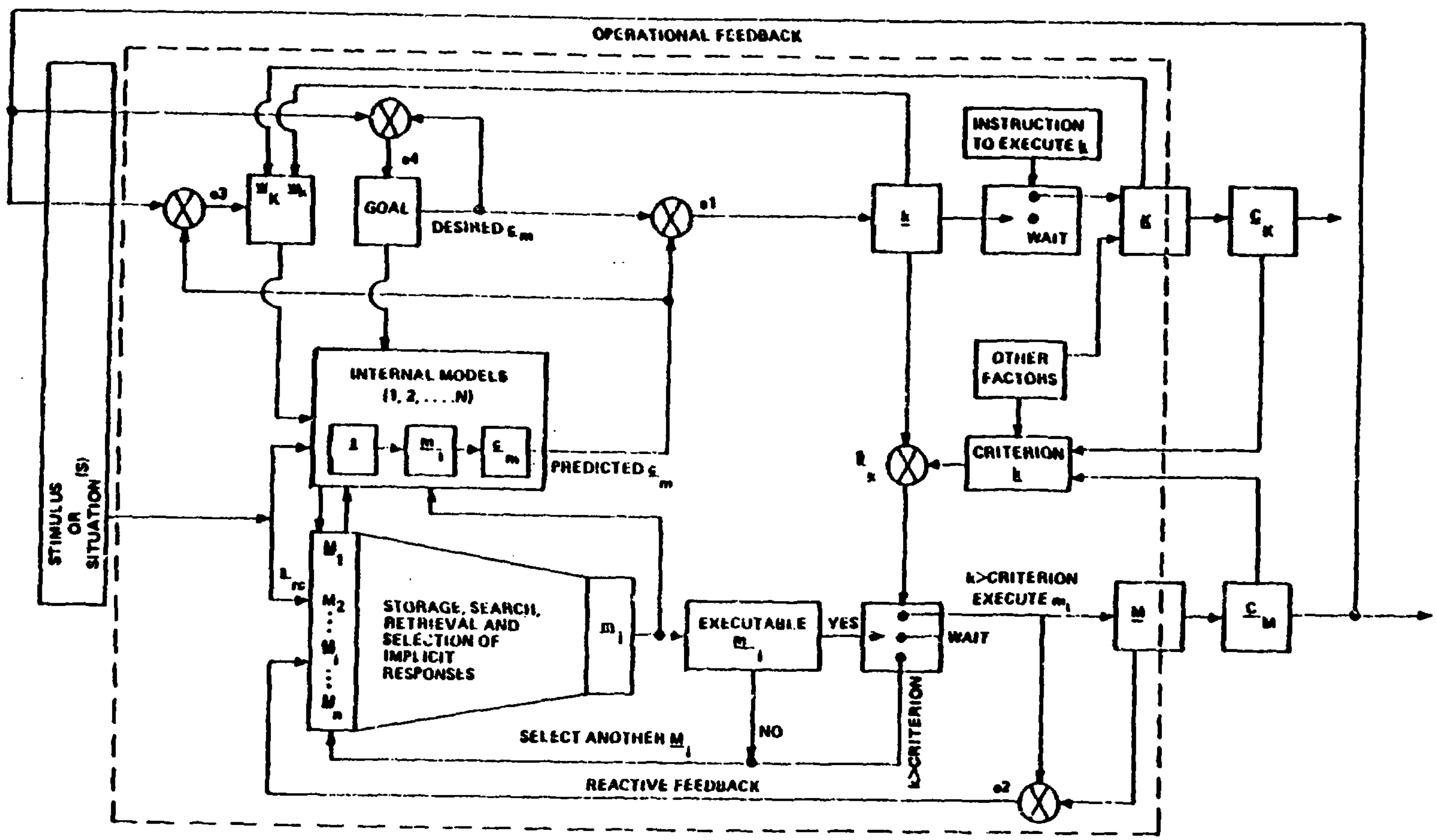

Figure 1. Conceptual framework of a human self-assessment process.

search/retrieval cycle is initiated. It is assumed that the speed of the decision to execute or to reject $\underline{m}_i$ is directly related to the difference between the level of $\underline{k}$ and the Criterion-$\underline{k}$, i.e., near threshold decisions take longer.

5. Actual consequences (C_M) are produced in the real world when a response (M) is executed. Information concerning these consequences is conveyed to the individual who utilizes this feedback information for at least two purposes:

   (a) The discrepancy between the actual C_M and the desired consequences influences the decision of whether to modify the Goal--or perhaps to continue responding to further reduce the discrepancy.

   (b) The discrepancy between the actual C_M and the predicted $\underline{c}_m$ permits a compensatory modification of the Internal Model to be made. The extent to which the Internal Model will be modified (for a given predictive discrepancy) is influenced by the covert sureness, $\underline{k}$, and the overt sureness, K, of the individual.

The main components of the SA process of interest in this report are discussed below.

The Goal

It is assumed that the responses or outputs of the individual are selected and executed for the purpose of attaining certain desired goals at any moment in time. Kelley (1968) points out that a typical feature of living organisms is the conception and choice among goals (p. viii). The notion that organisms behave in accordance with purposes is assumed by Miller, Galanter, and Pribram (1960). And Nuttin and Greenwald (1968) state, "the outcome of an action is regarded as playing a fundamental role in behavioral processes. Specifically, a future outcome can be said to determine behavior in the sense that the outcome is 'intended' prior to the performance of the action and the anticipation of the outcome subjectively appears to have the power of eliciting the action" (p. 2). In a relatively simple task such as paired associates learning it is assumed that the individual's goal is to be correct on each response.

The Internal Model

This is the process by which the individual is able to predict covertly the possible consequences or outcomes of various implicit responses which he may wish to test in fast-time. Attneave (1974) diagrams it in a "somewhat oversimplified way" (p. 494) as a stimulus-response-stimulus linkage:

$$S_1, R \to S_2$$

about which he says,

If situation S_1 obtains at a given time, and I do R, then situation S_2 results. If I know this, I know how to change situation S_1 into situation S_2. The beginning of knowledge, I think, is to be found in the fact that we live in a lawful world, in which propositions of this SRS type have some continuing validity from one day to the next" (p. 494).

This SRS view has been developed earlier in great detail by Tolman (1959) in relation to his theory of purposive behaviorism. The notion is that people develop cognitive models (or Internal Models) of the SRS kind which permit them to make covert predictions as to what the consequences would be if they were to execute some response. It is assumed that an adult individual, at least, possesses a fairly extensive repertoire of such Internal Models, from which he selects one (or more) depending upon the situation perceived to exist and the goals which are being sought. Presumably both the repertoire of models and the specifics of each internal model are developed through learning and experience in which the consequences of responses are predicted ($\underline{c}_m$) by the individual and then compared with the consequences which are produced (C_M) later when the selected response is executed.

This view seems consistent with Levine's (1975) characterization of adult human learning as the testing of hypotheses in a situation and the notion that learning involves searching for and finding the correct rule. Similarly Spear (1978) says that relationships "between events become stored as a memory together with specific attributes representing the context of those events" (p. 3). And Broadbent (1973) points out, "there is reasonable ground for believing that our brains calculate upon a model of the world the various consequences that will arise from different actions" (p. 180). Others (Miller et al., 1960) have used the term, "Image," to describe, "all the accumulated, organized knowledge that the organism has about itself and its world . . . (and) includes . . . his values as well as his facts" (p. 17). Recently Jagacinski and Miller (1978) stated, "It is a commonly accepted belief that humans use 'images' or internal models of the world around them in organizing and executing their everyday activities. The internal model concept is particularly prevalent in theories of decision making where actions are presumed to depend on the relationship between the individual's objectives and the anticipated results of his actions" (p. 425).

Bobrow's (1975) approach to the representation of knowledge within a (human or computer) system seems especially consistent with the above views and with the notion of an Internal Model. He proposes that representations (or Knowledge-states) result from a selective mapping of aspects of the real world. Thus, a Knowledge-state may be created which corresponds to a real World-state. Actions may be taken in the real world which alter the world from World-state-1 to World-state-2. If the world is altered, then some model operations exist in the system which make corresponding changes in the Knowledge-state from state-1 to Knowledge-state-2 (see Figure 2).

The manner in which particular real world actions are selected is not specified by Bobrow. However, he does point out that planning is a search for a series of action to bring about a particular desired world-state and

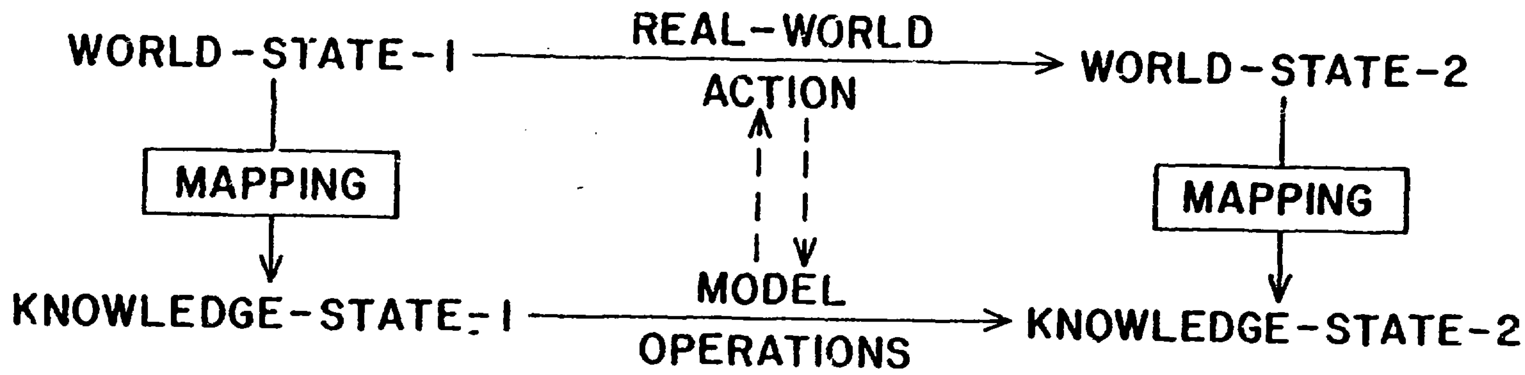

Figure 2. Diagram of changes in the world-state and in the knowledge-state by actions and model operations, respectively (After Bobrow, 1975).

says, "In planning, the changes are not real, they result from modeling activity, not world activity" (p. 12). Presumably, for a Knowledge-state-1 the system could enumerate a number of alternative model operations, then make estimations as to what Knowledge-state-2 would be produced by each operation; and finally select and translate one of the alternatives into real world action.

In Figure 2 the arrows going both ways between the Real-World Action and the Model Operations indicate that a person may develop his Internal Models through the observations that, if World-state-1 exists and if some Real-World Action occurs then World-state-2 will be produced. That is, a person can be a passive observer of S-R-S relationships and still develop these kinds of Internal Models. Indeed one would speculate that much of a person's knowledge of how to do things is acquired in this observational fashion.

Deese (1969) also seems to imply a similar internal process which he calls "understanding" which "only signals the potential for appropriate imagery, linguistic operations and other cognitive activity" (p. 516) and he indicates that people are capable of recognizing a state of understanding.

Bandura (1977) makes an important distinction between (a) outcome expectancies which are a person's estimates that given behavior will lead to particular outcomes and (b) efficacy expectancies which represent a person's convictions that he can successfully execute the behavior required to produce the outcomes. To the extent that the processes by which these two self-appraisals are made are different, this paper is concerned primarily with the outcome expectancies.

<u>Stimulus (S_1)</u>

This represents those stimuli which define what Attneave (1974) calls Situation 1, which includes the explicit stimuli which the experimenter presents on an experimental trial. This Stimulus performs two functions in the model:

a. It provides the input data to the person so that he can describe, to whatever extent he is able or is appropriate, the Situation 1 which prevails at the time. This is labelled "$\underline{s}$" in Figure 1. As Woodfield (1976) points out, "conditionals of the form 'If the environment were E_i, (the subject) would do B_i (where B_i is appropriate to (the Goal) in E_i' are true only on the assumption that if the environment were E_i, (the subject) would believe that it was E_i)" (p. 165).

b. It serves as a cue which initiates the memory search which in turn produces the retrieval or selection of an implicit response ($\underline{m_i}$)-- and influences the selection of the Internal Model. This selected response serves as an input to the ($\underline{s} \rightarrow \underline{m_i} \rightarrow \underline{c_m}$) Internal Model of the person which permits him to covertly assess the possible consequences relative to the goal. It seems likely that the efficacy expectancies proposed by Bandura (1977) are importantly involved in the process by which the implicit $\underline{m_i}$ responses are selected.

The retrieval cue property of the stimulus is labelled "$\underline{s}_{rc}$" in Figure 1. This label emphasizes the importance of the retrieval cue in the retrieval process as distinct from storage, memory, and forgetting. The distinction is of general importance in recall and recognition (Tulving, 1974; Rabinowitz et al., 1977; Broadbent & Broadbent, 1977), but it is of especial importance in considering the human self-assessment process.

A primary interest in the SA process in this paper is with the relation between (a) information which may, or may not, be stored in a person's memory and (b) the person's ability to validly and reliably determine that it is, or is not, stored in memory. The person's demonstration that some knowledge is possessed requires, in addition to its being stored, that it be retrieved under the circumstances which exist at the time of an inquiry or at the time when some utilization of the knowledge is necessary.

Covert Self-Assessment Response ($\underline{k}$)

The anticipated consequences, $\underline{c}_m$, of the selected $\underline{m}_i$ represent the output of the Internal Model. These anticipated consequences are compared with the desired consequences as specified by the Goal. The discrepancy or error is labelled "el." It is assumed that $\underline{k}$ is inversely related to the error 1, i.e., the greater the discrepancy then the lower the implicit sureness, $\underline{k}$. This $\underline{k}$ may be made observable by appropriately asking the person. However, there may be some distortions in the translation due to a number of factors. The $\underline{k}$ is assumed to serve two purposes.

a. It serves indirectly as a "gatekeeper" for the response $\underline{m}_i$ in the following way. The covert $\underline{m}_i$ will not be executed unless $\underline{k}$ attains some criterion level, as indicated by the comparison of $\underline{k}$ and the Criterion $\underline{k}$. It is further assumed that, once the Criterion $\underline{k}$ is attained, the vigor, speed, smoothness, etc. of M is directly related to $\underline{k}$. The greater is $\underline{k}$ then the more vigorously, quicker, smoother, etc. the M response is executed.

b. Also, $\underline{k}$ serves as a weighting factors ($\underline{w}_k$) in influencing the extent to which the Internal Model may be modified as a result of observing the actual consequences produced by the execution of a response. Some details of this modification are discussed later.

Regarding the gatekeeping function of $\underline{k}$ it may be noted in Figure 1 that when $\underline{m}_i$ is selected tentatively for testing then any one, but only one, of three things can happen: (1) $\underline{m}_i$ can be executed, (2) a search for another $\underline{m}_i$ can be initiated, or (3) $\underline{m}_i$ can be held in abeyance until either 1 or 2 is chosen. As stated earlier the speed with which a response is executed is assumed to be directly related to the extent to which $\underline{k}$ exceeds the criterion $\underline{k}$. Similarly, it is assumed that the latency of the rejection of an unacceptable $\underline{m}_i$ (and the initiation of a search for another $\underline{m}_i$) is related to the value of $\underline{k}$ such that the more sure the person is that the selected $\underline{m}_i$ is not an appropriate response, the quicker the search is re-initiated.

Consistent with this assumption is Kolers and Palef's (1976) finding of a general U-shaped function between the speed of responding and the frequency of occurrence of an item in the language. They presented, one at a time, 160 words which occur in language with high, medium, or low frequency and some nonwords--and asked subjects whether they knew the word well enough to be able to use it in a sentence. An analysis of the response latencies showed that "affirmations of negation were often more rapid than positive reports" (p. 553), i.e., subjects' responses that they did not know something were often faster than their responses that they knew something.

The findings of Murdock and Dufty (1972) are also consistent with this assumption. They found that the latency with which a visually presented item was recognized as having not been a member of a previously presented list (or as having been on the list) was inversely related to the confidence expressed (on a 6-point scale) by the subjects. They report that the responses of the subjects indicating the item had not been on the previous list were almost as fast as their responses indicating an item had been on the list.

Murdock and Dufty (1972) generally interpreted this finding as being consistent with the notion that the speed and confidence with which an item is recognized as being or not being a member of a previous list depends fundamentally upon the strength of the underlying memory trace rather than involving any separate process. If this interpretation is correct and sufficient then the proposal of a separate self-assessment process may be unnecessary.

However, Bernbach (1967) points out that a strength theory predicts that certain features of the receiver-operating-characteristic curves (which may be produced by a signal detection analysis of some learning data in which the learners have expressed a confidence in the correctness of each answer which they give) should be related to factors (such as the serial position of an item) which influence the strength of a response. He presents evidence which fails to support this prediction. Thus, it appears that even though strength theory alone may be quite adequate for the interpretation of recognition-memory data, it is not sufficient to account for people's confidence rating in some other kinds of learning situation. Bernbach describes a finite-state decision theory which is consistent with the evidence; and a separate self-assessment process such as is described in this present report may also be involved.

<u>Self-Assessments of Responses Which Are Called Either Correct or Wrong</u>. A situation which is conceptually awkward for the proposed model relative to <u>k</u> is one in which the response made is either "correct" or "wrong" as in a paired-associates learning task. A difficulty arises because there does not seem to be various degrees of discrepancy between the "goal" and the "predicted consequences."

However, the view that the response in the paired-associates learning task is either totally correct or wrong may obscure some relevant details. For example, for the person to make a correct response he must correctly accomplish a number of component subtasks or activities, e.g., he must detect and identify the stimulus, select a response, and execute the response

within the time limits. If any one of these components is deficient then the response is called "wrong." Thus, the person could correctly accomplish 90% of the components and still fail to make a "correct" response. Under the normal circumstances the consequences predicted by the Internal Model could reflect an accumulation of the components which are successfully accomplished, e.g., 90%. If the person's goal is to accomplish 100% of the components, then the notion that _k_ is inversely related to the discrepancy may be conveniently retained. This assumption that a learner's goal is to be correct all of the time is consistent with Sampson and Chen (1971) in their proposed model of human binary prediction behavior.

In a paired associates learning task, the subject is informed as to the correctness of his response. For example, after each response the subject may simply be told "correct" or "wrong"; or the stimulus may be presented along with the correct response, which permits the subject to infer the correctness of his response by comparing his recollection of the response which he just made with the presented correct response. In many experimental learning situations the subject will tend to repeat a response if it has previously been followed by "correct" and not repeat a response if it has been followed by "wrong."

Buchwald (1969) and others (d'Ydewalle & Eeleen, 1975) have proposed that the repetition of such a response which has been previously made depends upon whether the individual (a) recalls the response which was made previously and (b) recalls the consequences or feedback information relative to the previous response. In the Internal Model (Figure 1) these two recollections would refer to (a) the retrieval of the response $\underline{m_i}$ when Situation S_1 is presented and (b) the ability to predict the consequences $\underline{c_m}$ if $\underline{m_i}$ were to be made when Situation S_1 exists.

From this point of view and assuming that Situation S_1 is accurately perceived, the sureness _k_ would be a function of:

a. the probability that $\underline{m_i}$ will be retrieved and tested when Situation S_1, the Stimulus, is presented--which is equivalent to the probability of recalling the response that was previously made to the stimulus and

b. the probability that $\underline{c_m}$ will be recalled when $\underline{m_i}$ is tested in the Internal Model--which is equivalent to $p(\underline{c_m} \mid \underline{m_i})$ of the probability or recalling the previous consequences.

For illustration, consider a task in which one of two signal lights will be lit 4 to 5 seconds after the onset of a warning light; and the person's task is to predict, during the 4- to 5-second time period, which of the two lights will be lit. Let us, as the experimenters, arrange the circumstances so that Light 1 is lit on 80% of the occasions (at random) and Light 2 is lit on the other 20% of the occasions, i.e., $p\ (L_1) = 0.8$ and $p\ (L_2) = 0.2$.

In such a two-light prediction task, after a large number of trials, the relative frequency of the person's choice of Light 1 and Light 2, if no special reinforcements are delivered for correct responses, is typically found to be approximately 80% and 20%, respectively--called a matching choice strategy (Siegel, 1964).

As was stated earlier, assume that the person's Goal is to give correct answers all of the time. Presumably the person retrieves some response, say $\underline{m}_1$, and tests it in the Internal Model:

$$(\underline{s}) \rightarrow (\underline{m}) \rightarrow (\underline{c}_m = \text{correct 80\% of the time}).$$

If the learner's goal is to give correct answers all of the time then the discrepancy between the desired consequences (100% correct) and the predicted consequences (80% correct) is 20%; and the sureness in the correctness of the response may be relatively high, say 80%.

Model and Real-World Uncertainty. This line of thought indicates that $\underline{k}$ also depends upon the probabilistic relationships between $\underline{m}_i$ and $\underline{c}_m$ in the person's Internal Model. There are at least two major sources of this $\underline{m}_i \rightarrow \underline{c}_m$ uncertainty. First, the uncertainty could be due to the incomplete learning of the $[(\underline{m}_i \mid \underline{s}) \rightarrow \underline{c}_m]$ relation by the person; this might be called model uncertainty. Second, the uncertainty could be inherent in the real-world situation which the Internal Model represents; this might be called real-world uncertainty.

The two-light prediction task, in which the outcomes are probabilistically related to the responses, is an example of real-world uncertainty. It is expected that the amount of real-world uncertainty determines the limit of the sureness which the person may attain. In a two-light prediction task if the p (Light 1) is 0.8, then the maximum sureness an individual may properly attain for his choice of Light 1 is 80% because the real-world uncertainty is at that level.

On the other hand, in a typical paired-associates learning task, the $[(\underline{s}_1) \rightarrow (\underline{m}_1) \rightarrow (\text{correct})]$ relationship is fixed and, thus, the real-world uncertainty is virtually zero. An individual can reasonably be expected to attain a 100% sureness when the Internal Model is appropriately and completely developed.

In a two-light prediction task a person can be influenced to depart from a matching choice strategy toward a pure choice strategy of predicting the most frequently occurring event all of the time (Siegel, 1964). This may be accomplished by altering the experimental situation so that the person receives a payoff, say 25 cents, for making a correct prediction and a loss, say a loss of 25 cents, for making a wrong prediction.

It should be noted that the expected proportion of correct predictions for M_1 is 0.8 regardless of whether the person employs a matching or a pure strategy, i.e., approximately 80% of the M_1 responses will be correct regardless of how often M_1 is made. Similarly the expected proportion of correct predictions for M_2 is 0.2 regardless of how the person distributes his responses between M_1 and M_2 (provided M_1 and M_2 is made at all).

Thus, even though the relative frequency of choosing Light 1 may increase from 80% to near 100% when payoffs and losses are introduced the person's sureness in the correctness of the predictions would not be expected to increase in a comparable fashion. This is the case because the expected proportion of choices of one light (or the other) which is called correct is independent of the number of times the light is chosen.

Thus, the sureness which a person possesses regarding the correctness of his anticipated response or of his pre-feedback executed response depends upon not only the discrepancy between the goal and the predicted consequences, but also the probabilistic relations between $\underline{m}_i \rightarrow \underline{c}_m$ which exist at a particular time in the person's Internal Model.

Response Repertoire

At a molecular response level, this is quite similar to Adams' (1971) concept of a memory trace distribution, i.e., at a particular moment in time there is a variety of (simple) responses available with associated probabilities of being selected and initiated. At a more molar response level, one may think of the repertoire as being composed of a number of responses, sequences of responses, or possible plans of action which might be executed by the person (Miller et al., 1960; Attneave, 1974).

Another feature of the response repertoire needs to be mentioned. Once the implicit response, $\underline{m}_i$, is initiated then its overt execution, M, is monitored through a reactive feedback loop so that a suitable fidelity between the specification of $\underline{m}_i$ and the execution of M is maintained (Figure 1). Having "Error 2" enter the $\underline{m}_i$ response repertoire is intended to suggest that the execution of a particular M produces some modification of that response repertoire.

It is reasonable to suppose that the location of the feedback loop (extrinsic, intrinsic, or central) as well as the locus of the standard may depend upon the hierarchical level of the response. For example, there is some evidence (Roy & Marteniuk, 1974) that simple motor responses of, say, less than 150 msec. are controlled by different loops than are similar responses of 1 sec. or longer. Indeed, the satisfactory execution of some selected responses may not require such closed loop control involving sensory feedback at all. For example, Kelso (1977) suggests that certain simple psychomotor responses, such as blindly positioning the index finger to a position which the person has previously defined by his own movement, depends only upon the availability of a control movement plan to guide it and not upon the feedback of response-produced sensory information.

It is assumed that there may be stored in memory an extensive repertoire of Internal Models of an $[(\underline{s}) \rightarrow (\underline{m}_i) \rightarrow (\underline{c}_m)]$ kind. A person retrieves a specific Internal Model based upon his analysis, $\underline{s}$, of the situation and upon the Goal. It is this specific Internal Model into which a selected $\underline{m}_i$ is inserted to anticipate the consequences, $\underline{c}_m$, of the response.

Thus, one's sureness or self-assessment response may be inaccurate because an inappropriate Internal Model is used for the anticipation of the consequences of a response. Woodfield (1976) states, "it is not always the case that if $\underline{S}$ (the subject) correctly believes that the situation is E_1, $\underline{S}$ performs the response which is, in fact, a means to (the goal) in E_1. $\underline{S}$ may be right about the situation, but wrong about the best way to get to (the goal) in that situation. $\underline{S}$ does what he believes to be appropriate" (p. 165).

Operational Feedback

The overt response, M, produces consequences, $\underline{C}_M$, in the real-world. Information concerning these actual consequences is often received by the responder. This feedback information allows a comparison to be made between (a) the perceived real-world consequences and (b) the earlier predicted consequences. Any perceived discrepancy between C_M and $\underline{c}_m$, which is labelled e3 (Error 3) in Figure 1, may result in a compensatory modification of the Internal Model. This modification is part of what may be called learning or increasing one's knowledge. As Bandura (1977) points out, "Learning from response consequences is . . . conceived of largely as a cognitive process. Consequences serve as an unarticulated way of informing performers what they must do to gain beneficial outcomes and to avoid punishing ones" (p. 192).

The covert and overt self-assessment responses, $\underline{k}$ and K, are hypothesized to play an important role in the operational feedback loop. They serve to weight ($\underline{w}_k$ and $\underline{w}_K$) the influence of Error 3. For a given size of discrepancy between the predicted and actual consequences of a response, the extent to which the Internal Model will be modified is affected by the sureness which the individual possessed regarding the response prior to its execution. One might speculate that the stronger the belief, i.e., the higher the sureness in the correctness of the response, then the more resistant are the components of the Internal Model to being modified by a disconfirmation.

Shuford et al. (1967) have distinguished between uninformed and misinformed individuals. The difference between a misinformed and uninformed individual may be represented as shown below. In terms of the Internal Model a misinformed state would suggest that the association, $(m_i \mid \underline{s}) \rightarrow \underline{c}_m$ is well established but wrong (although there may also be a misperception of the Situation).

	Sureness of correctness	
$\underline{M}$ - Response	Unsure	Sure
Correct	Uninformed	Informed
Wrong	Uninformed	Misinformed

If a misinformed state is reflected by a high sureness in a wrong or inappropriate response and if the extent to which the feedback produces a modification of the Internal Model is influenced by the sureness, as described earlier, then it would be of especial interest to observe the changes which take place in such misinformed, in contrast to uninformed, responses with practice and successive disconfirmation.

Consequences of the Overt Self-Assessment Response ($\underline{C}_K$)

In most human learning or performance research, no overt self-assessment response is required of subjects. In the relatively few studies in which K has been required, the consequences, C_K, associated with the self-assessment responses have not been experimentally manipulated. Based upon the preliminary

model it is hypothesized that the C_K will affect certain characteristics of human performance--presumably by affecting the Criterion $\underline{k}$. However, the accuracy with which the comparison between the desired $\underline{c}_m$ and the predicted $\underline{c}_m$ is accomplished may also be affected by C_K.

Goal Modification

Central to Kelley's (1968) discussion of the human as a component in the control process is the notion that people are able to conceptualize possible goals and to choose from among them. Given that some goal exists or is conceptualized and chosen at one moment in time, it is subject to being modified. Among the variables which may affect the modification of the goal is a discrepancy between (a) the state of affairs conceptualized by the individual as the goal and (b) the actual consequences perceived to exist following the execution of some response(s). In Figure 1 the result of this comparison is labelled e4 (Error 4). Based upon this comparison the person may modify his goals.

Miller et al. (1960) seem to view the goal modification in a similar way when they say, "An alternative to the stop-rule (for searching) is a modification of the conditions that are imposed in the test phase. After searching unsuccessfully for a pen, we settle for a pencil" (p. 171). Thus, a goal may be modified prior to the overt execution of a response as well as after a comparison has been made between the consequences of an executed response and the desired state of affairs.

AN EXPERIMENT

It would be premature to decide now whether it is necessary or even desirable to employ a notion of a "self-assessment" process, as outlined above, which is separate from concepts already available to explain and predict the ability of people to express various levels of sureness in the correctness of responses which they anticipate making or have already made. For example, an appropriate use of the enduring concept of associative strength may be sufficient to account for the self-assessment responding which is of interest in this paper. At present, however, an interpretation of self-assessment responding based upon associative strength theory alone seems incomplete (Bernbach, 1967). To permit existing concepts to be refined and choices among concepts to be made it will be helpful to collect additional data relevant to the self-assessment process, generally, and to the proposed model, specifically.

First is a general question of whether learning of new responses is affected by the concomitant performance of a self-assessment task, i.e., is the rate at which behavior is modified by practice either retarded or expedited by the performance of a secondary task of self-assessment. For example, the additional information processing and other associated responses demanded by the performance of the self-assessment task while a person is engaged in acquiring new responses may interfere with the primary task of learning.

On the other hand, the performance of a self-assessment task might expedite learning. For example, the self-assessment task may require the

learner to attend more closely to various features of a stimulus, a response, the response consequences, or the relations among them; or the execution of a self-assessment response may provide an extra source of reinforcement, e.g., making an accurate self-assessment along with making a correct response to-be-learned may be more rewarding than simply making a correct response along. Kanfer (1971) suggests that requiring people to attend to their own actions and to the effects of their actions may have reactive effects such as "modifying the very behavior which they are intended to describe" (p. 56). Also, Wade (1974) as well as some preliminary work by the author suggests that acquisition may benefit from the simultaneous performance of a secondary self-assessment task.

To explore this hypothesis different groups of subjects were given a primary task of paired-associates learning. Groups of subjects who performed a self-assessment task, using either two, four, or eight K-response categories, were compared with control groups who either performed only the primary learning task or, in addition to that, made simple motor responses instead of self-assessment responses.

A second hypothesis, closely related to the first, is that the extent to which the probability of occurrence of the M response is modified depends not only upon whether the executed M response is perceived by the learner as being correct, but also upon the sureness which the individual possesses (and indicates) about the correctness of the response. The model of the self-assessment process proposes that the covert $\underline{k}$ and overt K responses serve to weight the effects of feedback information or knowledge of results.

For example, an M response about which a person is sure of its correctness may be relatively resistant to modification. This could be expected because the level of sureness may be seen as a reflection of strength of association between the stimulus and the M response or between the M response and the consequences of the response which are predicted.

However, if the self-assessment responses are interpreted as potential sources of reinforcement, then the confirmation or disconfirmation of the K response must be considered. For example, a wrong response about which the learner is unsure should show more resistance to modification than a wrong response about which he is sure because the presentation of the knowledge of results confirms the K response of unsure. On the other hand the observation of a relative persistence of wrong M responses about which the learner has indicated a high level of sureness in their correctness would be consistent with an associative strength interpretation. The tenability of these interpretations is tested by requiring the paired associates learner to indicate a level of sureness when each M response is made.

A third hypothesis of interest concerns the order in which the $\underline{m}$ response and the $\underline{k}$ response are internally or covertly selected by the person. The model indicates that, first, an $\underline{m}$ response is retrieved and tested which, in turn, produces a level of sureness, $\underline{k}$. This hypothesis is tested by requiring half of the learners to execute the M response first, followed by a K response; the other half of the learners are required to execute the M and K responses in the reverse order.

The speed with which the execution of the two-response sequence is completed should reflect the order in which the m and k responses are internally processed and selected. The KM order of response execution should exhibit the shortest total time because the MK subjects require the retrieval of a previously selected k response which the KM subjects do not require. This assumes that both m and k responses are selected before either is executed.

Finally, it is of interest to determine the accuracy with which people who are engaged in a learning task can assess the correctness of a response which they will later execute--or have already executed but have not yet received any extrinsic feedback or knowledge of results about its correctness.

Method

Subjects. Ninety female and 90 male students served as subjects as part of their requirements in an introductory psychology course at New Mexico State University.

Primary Learning Task. The primary task of all subjects was to learn the correct names of eight different pairs of hand pliers. Drawings of eight hand pliers were constructed based upon a review of the pictures of pliers contained in military tool catalogues. The drawings were composed by combining two different plier heads (a short broad head and a long slender head) with two handle shapes (symmetrically curved and nonsymmetrically curved) and the handles were either cushioned or uncushioned. These eight (2 x 2 x 2) different pictures of pliers served as the stimuli for a paired associates learning task.

The response terms of SHAPE, BEND, FORM, and TWIST were initially assigned randomly to the four long slender headed pliers and the terms SPLIT, CUT, CLIP and SNIP were assigned randomly to the four short broad headed pliers. Once assigned these names were the same for all subjects in all conditions throughout the experiment. The subjects indicated their answer by pressing one of eight labelled buttons on a response panel. The eight verbal response terms were nonsystematically assigned as labels to the eight buttons and were, from left to right, TWIST, CLIP, FORM, SPLIT, SHAPE, SNIP, BEND, and CUT. Once assigned they were the same for all 180 subjects.

Apparatus. A teletype permitted commands to be given to a PDP-8E computer, to print data, and, also, punch data on paper tape. The computer controlled the presentation of (a) an easily heard tone through Telex 1210-02 earphones, (b) the stimulus pictures for 8 seconds, and (c) a knowledge-of-results slide, after a 1.5-second delay, which contained the stimulus picture along with its correct name for 4 seconds. The stimuli and knowledge-of-results slides were rear projected on a 11.4 cm. x 12.7 cm. screen. The subject's viewing distance was approximately 60 cm.

A 76.4 cm. x 47.5 cm. response panel was laterally centered 29.5 cm. below the center of the projection screen. The panel was generally horizontal but the front edge was tilted approximately 20 degrees downward to be more normal to the subject's line of vision and to be more convenient

manually. Internally lit 2.8 cm. wide x 2.1 cm. long response buttons on the panel could be arranged in nine different ways (see Figure 3) each corresponding with one of the nine different treatments involved in this experiment. For all button arrangements a START button was centered laterally and 19.3 cm. from the front edge of the response panel. The rows of buttons were separated 7.5 cm. on center. The buttons were laterally separated 1.0 cm. edge-to-edge, except the buttons next to the center line which were separated by 4.8 cm.

A timer measured to the nearest millisecond the latencies from the onset of the stimulus to the release and activation of the various buttons. Each answer response, its correctness, and each K response were recorded for each stimulus presentation.

Procedures. The primary task of the subject, which was to learn the names of the pliers, was considered accomplished when the subject could go through the list twice consecutively with no error. It was necessary for the subject to have the START button depressed at the time the stimulus was about to be presented; a 100 msec. tone was presented 1.5 sec. prior to the stimulus presentation. If the subject did not have the START button depressed at the time when the tone was presented, then the tone remained on until either the START button was depressed or 1.5 sec. had elapsed, whichever occurred first.

Ten females and 10 males were nonsystematically assigned to each of the following 9 treatments:

M: The subjects performed only the primary learning task, making an answer (M) response in Row 2, the row of buttons immediately above the START button (see Figure 3).

MX: First made an M response in Row 2 followed by pressing a single button (X) which was labelled "RECORD" in Row 3.

MK2: Made an M response in Row 2 followed by a self-assessment (K) response in Row 3. Two K-response categories were available. The button on the left end of Row 3 was labelled "NOT SURE" and the button on the right end of Row 3 was labelled "SURE."

MK4: Made an M response in Row 2 followed by a K response in Row 3. Four K-response categories were available. The buttons on the left and right ends of Row 3 were labelled the same as treatment MK2. Each of the two intermediate buttons had a 2mm x 2mm black square in its center.

MK8: Made an M response in Row 2 followed by a K response in Row 3. Eight K-response categories were available. The buttons on the left and right ends of Row 3 were labelled the same as treatments MK2 and MK4. Each of the six intermediate buttons had a 2mm x 2mm black square in its center.

XM: First press a single button which was labelled "RECORD" in Row 2 followed by an M response in Row 3.

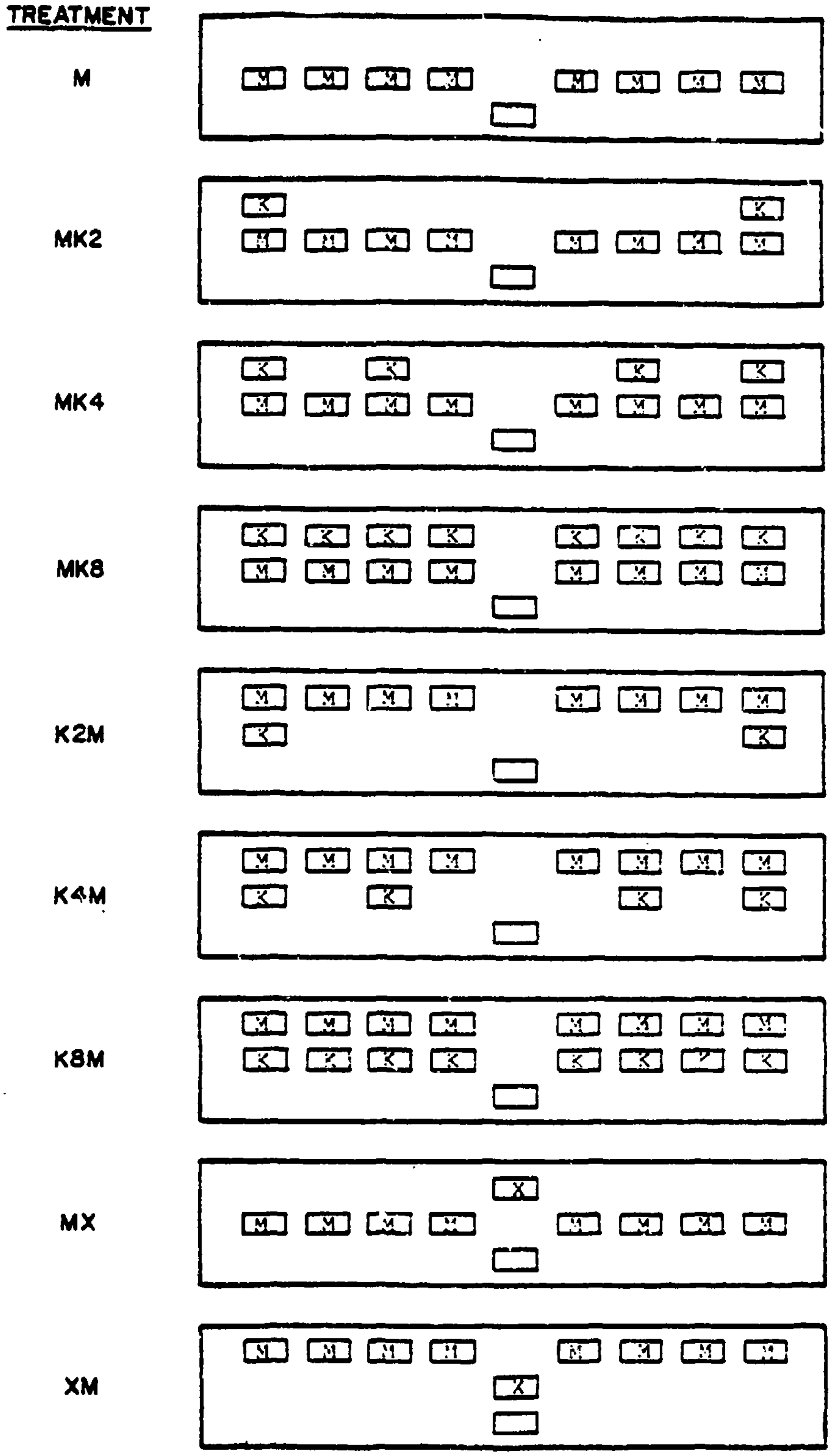

Figure 3. Button arrangements on the panel for nine different treatments.

K2M: First made a K response in Row 2 followed by an M response in Row 3. Two K-response buttons were available in Row 2 with the same labels as used for treatment MK2.

K4M: First made a K response in Row 2 followed by an M response in Row 3. Four K-response buttons were available in Row 2 with the same labels as used for treatment MK4.

K8M: First made a K response in Row 2 followed by an M response in Row 3. Eight K-response buttons were available in Row 2 with the same labels as used for treatment MK8.

Five male and five female subjects were treated under each of the nine treatments by each of two (male) graduate student experimenters. Subjects were tested at five specific times of day (1115, 1300, 1430, 1600, and 1730) and Mondays through Fridays. An attempt was made to test four subjects under each treatment on each day of the week. Although this was not possible considering other counterbalancing requirements, we did test at least three and no more than five subjects under each treatment on each of the five days of the week such that a total of 40, 35, 35, 35, and 35 subjects were tested on Mondays, Tuesday, . . . and Fridays, respectively. Similarly, approximately an equal number of subjects were tested by each experimenter at each of the five times-of-day under each of the nine treatments.

After the subject was seated, instructions appropriate for the assigned treatment were read which (a) stated that the task was to learn the names of eight different pliers, (b) pointed out the critical differences in the plier heads, handle shapes, and cushioning using enlarged (19 cm wide x 9 cm) pictures of the stimuli, (c) informed the subject of the various buttons on the panel, their functions for the treatment under which he was being tested, and that the buttons should be pressed as quickly and as accurately as possible, (d) informed the subject of other details, e.g., that the START button must be depressed by the time scheduled for the stimulus presentation and that 8 seconds were available to press the answer button and, if the assigned treatment required, press a self-assessment button, and (e) informed subjects that they should press the buttons always using one and the same finger and inquired which finger they would use (173 said right index, 1 right middle, and 6 left index).

Immediately after the instructions were read to the subject, the eight stimulus pictures of the pliers along with their correct response names were projected one time each for 5 seconds. Then the learning session immediately began with the projection of a single small circular black dot for 3 seconds; the warning tone sounded, the subject depressed the START button, and the stimuli were presented one at a time for 8 seconds, and the subject was required to respond as indicated by the assigned treatment.

At the end of the series of eight stimuli there was a 5½-second delay, then the dot slide was presented, and the eight stimuli were again presented one at a time, but in a different order. At the end of the first two trials of eight stimuli each, the experimenter interrupted the session for 1½ minutes during which he remedied (by paraphrasing appropriate parts of the initial instructions) any procedural difficulties which

the subject appeared to be having and emphasized that it was important for the subject to make all responses even though she/he wasn't sure of them.

Then the session was resumed. Seven different orders of stimuli were used and these orders were recycled through until the subjects had learned the names. If the names of the pliers had not been learned by the end of the 40th trial then the session was ended, the data of the subject were declared unacceptable, and another substitute subject was tested on a subsequent week under the same conditions (treatment, sex, day of week, time of day, and experimenter).

Results and Discussion

Self-Assessment and Acquisition. The mean number of trials required to attain three different levels of acquisition (50%, 75%, and 100% correct) under each of the nine treatments is shown in Figure 4. The operational definitions of the three acquisition levels are:

Low: The first series of eight stimuli (called a trial) on which four (50%) or more of the eight answers were correct and at least two correct answers occurred on each subsequent trial.

Medium: The first trial on which six (75%) or more correct answers were made and at least four correct answers occurred on each subsequent trial.

High: The first of two consecutive trials on which no error (100% correct) was made.

Dunnett's test (Winer, 1971, p. 201) comparing group M with groups MK2, MK4, and MK8 showed that group MK8 required significantly fewer trials to attain the low, medium, and high acquisition criteria, $tD(76, 4) = 2.88$, 3.00, and 5.66, $p < .05$, respectively, than did group M; groups MK2 and MK4 were significantly different from group M only at the high acquisition level, $tD(76, 4) = 3.16$ and 5.17, respectively, $p < .05$. A Dunnett's test involving the KM groups showed that only groups K8M and K4M at the high acquisition level are reliably different from group M, $tD(76, 4) = 3.57$ and 2.67, $p < .05$, respectively.

An analysis of variance of these data for the six treatments which required the subjects to make self-assessments showed that the main effects of the order (O) in which the answer and self-assessment responses were executed was significant, $F(1, 96) = 4.08$, $p < .05$. It may be seen in Figure 4 that more rapid acquisition is associated with those groups of subjects who indicated their sureness after giving their answer. Inspection of Figure 4 also suggests that as practice proceeds the relative beneficial effects of the MK order of responding becomes greater; and this is statistically supported, $F(2, 192) = 3.11$, $p < .05$.

A more detailed statistical analysis revealed that a specific effect of the MK treatment on the number of trials required to learn the material was to reduce the upper extreme scores. For example, 8 of the 20 subjects under treatment M required more than 21 trials (the third quartile

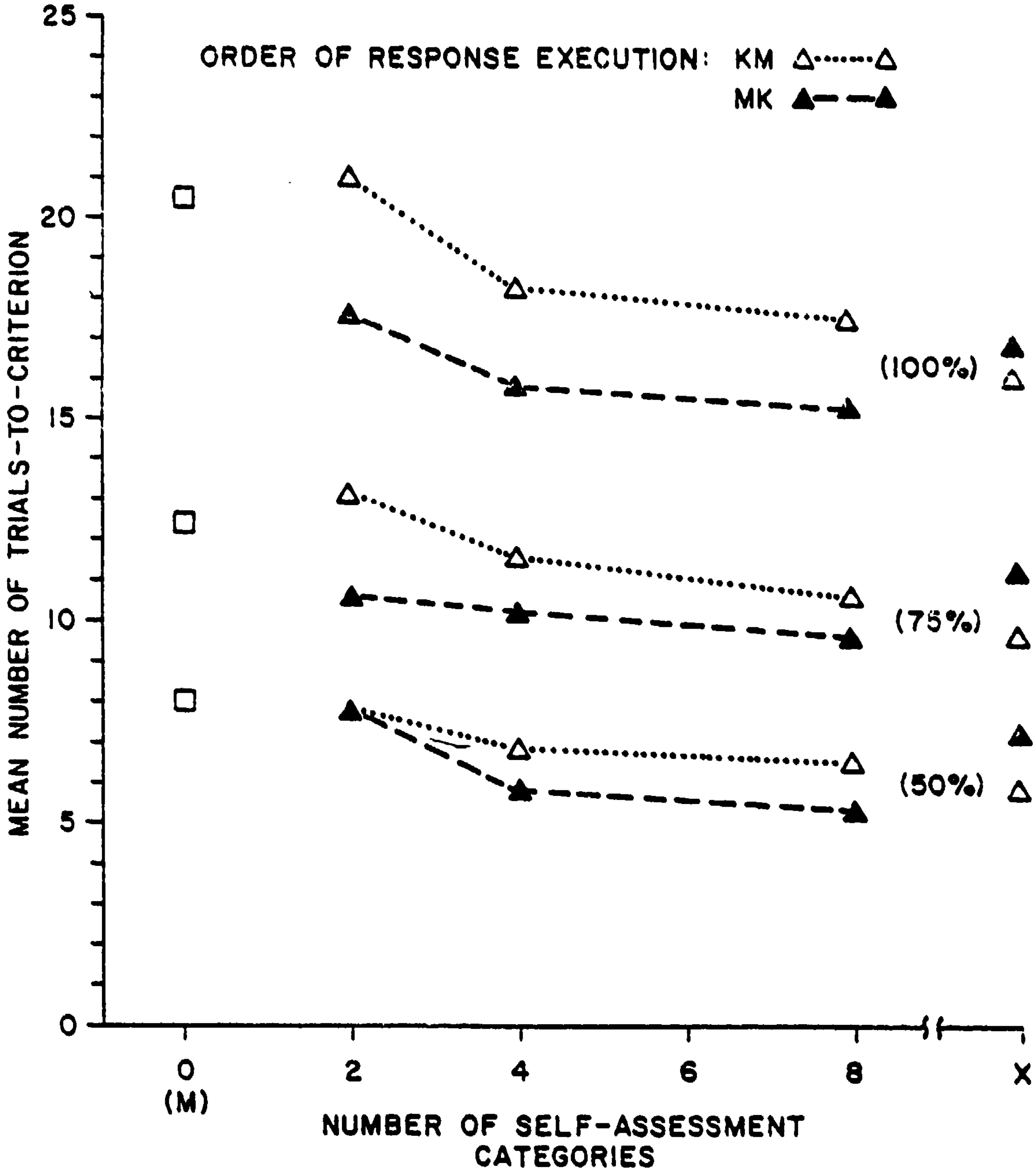

Figure 4. Mean number of trials required to attain the 50%, 75% and 100% criterion of correctness as a function of the number (2, 4 or 8) of self-assessment response categories used by the learner and for the control groups (M, MX, and XM). Separate plots are shown for those groups of subjects who executed their answer first (MK) and for those who executed their self-assessment response first (KM). Each mean is based on twenty subjects.

of the combined treatments) to learn the material while no subject in MK8 required more than 21 trials. Based upon the principles of the Brown-Mood median test (Bradley, 1968), differences among the M and MK treatments at the high criterion level in terms of the number of subjects above the third quartile was significant, $\chi^2(3) = 10.14$, $\underline{p} < .05$. These differences among treatments were not significant at either the median or the first quartile, $\chi^2(3) = 2.11$, $\underline{p} > .10$, and $\chi^2(3) = 3.73$, $\underline{p} > .10$, respectively.

This suggests that the effects of self-assessment responding on learning may be different for different people, e.g., those learners who would normally require an extremely large number of trials to learn the material benefit more from the effect of self-assessing than those learners who already would learn the material rapidly. However, a "floor" effect which limits the smallest number of trials necessary to learn the material may prevent the effects of self-assessment on acquisition from being observed for the more rapid learners. Some other measures of learning, e.g., measures of retention, may be more sensitive to the effects of self-assessment.

Subjects in the MK and KM groups were required to determine their level of sureness and to indicate the level by the execution of a motor response, i.e., pressing a button. Groups MX and XM were included in the experiment to identify the effects of the extra motor component associated with the execution of the SA response separate from the cognitive self-assessment component. Dunnett's test showed that both the MX and XM groups learned the material to the high criterion in fewer trials than the control group M, $tD(57, 4) = 4.16$ and 5.13, $\underline{p} < .05$, respectively; also group XM is significantly different from group M at the medium acquisition level, $tD(57, 4) = 3.25$, $\underline{p} < .05$. Similar statistical comparisons showed that neither the MK nor the KM groups learned the material in reliably fewer trials than their respective MX and XM control groups.

The findings concerning the MX and XM motor-control groups are difficult to interpret. All groups pressed buttons to indicate each answer. The extra motor activity associated with the execution of the single (X) response is relatively modest. Thus, it does not seem reasonable to attribute the more rapid learning by the MX and XM groups to that extra motor activity. A possibility is that the requirement to execute a sequence of two responses (which is the case for all groups except group M) demands that some response program be developed; and this development of a two-response program may be accompanied by a greater amount of covert rehearsal of the response to-be-learned.

Taken together these data suggest that requiring learners to perform the SA task may expedite acquisition (at least for those learners who otherwise would require an extremely large number of trials) especially if the SA response is executed after the response to-be-learned is given. However, the motor component may play a more complex role in learning than anticipated and the extent to which the actual self-assessment is sufficient or necessary to expedite acquisition is not clear.

In terms of applying these findings to training situations it is of interest to note that there is no hint in the data that any of the components, e.g., motor components or additional information processing, involved in the performance of the SA task interferes with the primary

learning in this situation. More detailed interpretations are offered later in the general discussion.

Specific Effects of Self-Assessments on Response Modification. The second hypothesis is that the extent to which a specific response will be modified with practice depends upon the sureness associated with its execution. Figure 5 shows the mean proportion of the total responses which fall into each self-assessment category for correct and wrong responses at the low and medium acquisition level under treatments MK2, K2M, MK4, K4M, MK8, and K8M.

An inspection of these figures suggests that there would be little risk of misinterpreting these results if all of those K responses other than "Sure" were lumped together and called "Unsure" self-assessments. In this article "Sure" and "Unsure" always refer to whether the subject is sure or unsure that the response is correct.

The proportion of sure and unsure responses for the various treatments is shown in Table 1 separately for correct and wrong responses at the low and medium acquisition levels. The statistical reliability of the differences between the proportions at the two acquisition levels is also indicated in Table 1.

It can be seen that for all treatments the increase with practice in the proportion of sure-correct responses and the decrease in the proportion of unsure-wrong responses is statistically significant. The change with practice in the proportion of unsure-correct and of sure-wrong responses is not significant for any treatment.

Thus, it seems that the changes in responses which occur with practice (during the middle stages of acquisition at least) involve a shift in the proportion from unsure-wrong responses to sure-correct responses, with little or no change in the proportion of sure-wrong or unsure-correct responses. Table 2 shows the decreases in the proportion of unsure-wrong responses and increases in the proportion of sure-correct responses which occurred with practice from the low to medium acquisition level. None of the differences between the increases and decreases is statistically significant. Thus, over the middle stages of practice it is only the accurately self-assessed responses (sure-correct and unsure-wrong) which exhibit a change in their relative frequency of occurrence; the proportion of inaccurately self-assessed responses remains unchanged.

The relative constancy with practice of the proportion of sure-wrong responses is consistent with the notion that wrong responses about which the learner is sure of their correctness are especially resistant to change. However, the notion of persistence would also require that the specific responses which were sure and wrong be repeated relatively more often compared to, say, the unsure-wrong responses.

Table 3 shows the extent to which kinds of responses (sure-correct, unsure-correct, sure-wrong, and unsure-wrong) made on the low acquisition trial remain of the same kind or change to other kinds of responses on the medium acquisition trial. It may seem that only 14% of the sure-wrong responses on the low acquisition level were repeated as sure-wrong responses

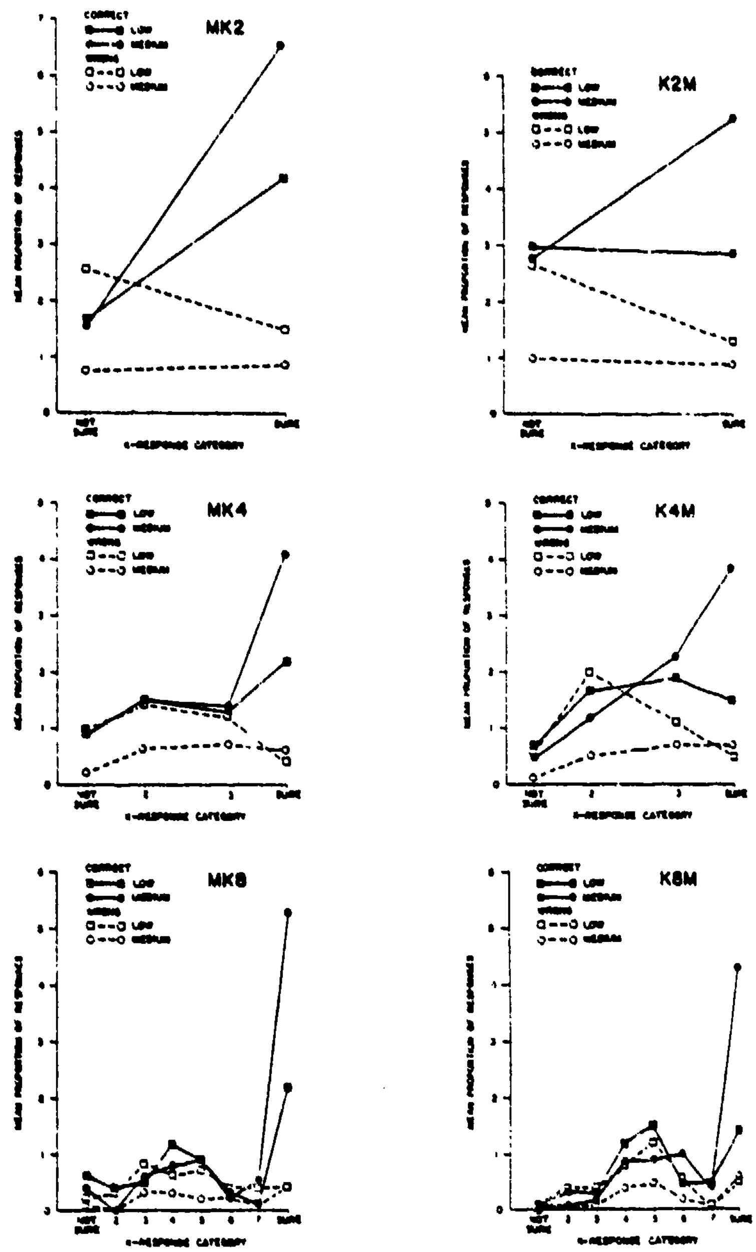

Figure 5. Mean proportion of the self-assessment (K) responses placed in each of the K-response categories for correct and wrong responses at the 50% and 75% criterion of correctness. Separate plots are shown for treatments MK2, K2M, MK4, K4M, MK8, and K8M.

Table 1

Mean Proportion of the Total Number of Responses Which Were Sure or Unsure and Correct or Wrong Under Each of the Six Treatments Involving Self-Assessments at Two Levels of Acquisition (Low and Medium)

	Correct				Wrong			
	Low		Medium		Low		Medium	
Treatment	Unsure	Sure	Unsure	Sure	Unsure	Sure	Unsure	Sure
MK2	.17	.42	$.16^{ns}$	.66*	.26	.15	.08*	$.09^{ns}$
K2M	.31	.29	$.28^{ns}$	.53*	.27	.13	.10*	$.09^{ns}$
MK4	.37	.22	$.38^{ns}$	.41*	.36	.04	.15*	$.06^{ns}$
K4M	.43	.16	$.40^{ns}$	.39*	.37	.05	.13*	$.07^{ns}$
MK8	.40	.22	$.34^{ns}$	.53*	.34	.04	.12*	$.04^{ns}$
K8M	.44	.14	$.36^{ns}$	.43*	.37	.05	.14*	$.06^{ns}$

^{ns}The difference between the low acquisition level and medium acquisition level is not statistically significant.

*The difference between the two acquisition levels is statistically significant at $p < .001$.

Table 2

Decrease in Proportion of Unsure-Wrong Responses and Increase in Sure-Correct Responses with Practice

Treatment	Decrease in Unsure-Wrong Responses	Increase in Sure-Correct Responses
MK2	.18	.24
K2M	.17	.24
MK4	.21	.19
K4M	.24	.24
MK8	.22	.31
K8M	.22	.29

on the medium acquisition level; while 22% of the unsure-wrong responses persisted as unsure-wrong responses. This difference (22% vs. 14%) is in the wrong direction to support the view that sure-wrong responses are resistant or insensitive to disconfirmations.

Table 3

Mean Proportion of Kinds of Responses (Sure-Correct, Unsure-Correct, Sure-Wrong and Unsure-Wrong) on the Low Acquisition Trial Which Persisted or Changed to Other Kinds of Responses on the Medium Acquisition Trial

Responses on Low Acquisition Trial		Responses on Medium Acquisition Trial			
		Sure		Unsure	
		Correct	Wrong	Correct	Wrong
Sure	Correct	.85	.08	.05	.02
	Wrong	.69	.14	.15	.03
Unsure	Correct	.41	.02	.49	.07
	Wrong	.32	.09	.37	.22

Indeed it seems that wrong responses about which a learner is sure of their correctness may be less likely to be wrong on the subsequent criterion trial (17%) than are wrong responses about which the learner is unsure (31%), $z = 2.54$, $p < .01$. From a reinforcement point of view one might expect unsure-wrong responses to be repeated more often than wrong responses about which one is sure of their correctness. The notion, mentioned earlier, is that reinforcement may be associated with the accuracy of a self-assessment response as well as with the correctness of the answer response; and the probability of an answer response being repeated is increased or decreased as a result of the two possible reinforcement events: the correctness of a response and the accuracy of a self-assessment.

The proportion of times in which a response which was both correct and accurately assessed on the low acquisition trial was also both correct and accurate on the medium acquisition trial, or a wrong and accurate response remained wrong and accurate, etc., is shown in Table 4.

The proportions in Table 4 are in accordance with a reinforcement interpretation. Those responses which were correct were repeated more often (.67) than those responses which were wrong (.18), $z = 7.60$, $p < .01$. And those responses which were accurately self-assessed were repeated more often (.53) than those responses which were inaccurately self-assessed (.31), $z = 3.45$, $p < .01$.

Table 4

Proportion of Different Kinds of Responses Made on the Low Acquisition Trial Which Were Repeated on Medium Acquisition Trial, e.g., a Response Which Was Correct and Inaccurately Assessed, i.e., Unsure on Low Acquisition Trial, Remained Correct and Inaccurate on the Medium Acquisition Trial

	Answer (M) Response		
Self-Assessment (K) Response	Correct	Wrong	Mean
Accurate	.85	.22	.53
Inaccurate	.59	.14	.31
Mean	.67	.18	

Of special interest is the apparently higher repetition (.22) of the accurately assessed, but wrong, responses, i.e., wrong responses about which the learner was unsure of the correctness, than of the inaccurately assessed wrong responses (.14), i.e., wrong responses about which the learner was sure of their correctness. However, the difference between the 22% repetition of accurate wrong responses and 14% repetition of inaccurate wrong responses is not statistically reliable; $z = 1.61$, $p > .05$. Thus, it seems that the effect of self-assessments on response repetitions may be restricted to correct responses.

Covert Selection of the Answer, m, and Self-Assessment, k, Responses. The third experimental hypothesis concerns the order in which the m and the k responses are internally selected. Is the m response selected first followed by a self-assessment of its correctness? Or does the self-assessment play such an intimate role in the selection of the m response in this learning task that k is already available by the time the m response has been selected?

If the m response is selected first and if it is assumed that both responses are selected before either is executed, then the KM treatments should exhibit a shorter response latency than the MK treatments because the MK treatments involve an additional step of retrieving a previously selected k response which the KM treatments do not require.

The mean response latencies (measured from the onset of the stimulus to the completion of the M-K or K-M response sequence) for the six self-assessment treatments at three levels (L) of acquisition are presented in Table 5.

An analysis of these latencies shows that the effect of L is significant, $F(2, 216) = 287$, $p < .001$; and the L interacts significantly with both the order (O) in which the answers and self-assessment responses were executed, $F(2, 216) = 4.20$, $p < .05$, and with the number of self-assessment categories (A), $F(4, 216) = 4.09$, $p < .01$. The main effect of neither O

nor A is statistically significant. At the low acquisition level the KM groups respond approximately 200 msec. faster than the MK groups. As practice proceeds the effects of the order in which the responses were executed vanishes.

Table 5

Mean Response Latencies, in Seconds, from Stimulus Onset to Initiation of Second Response for the Six Treatments Which Required Learners to Assess the Correctness of Their Own Responses (Either Before, KM, or After, MK, Each Answer) at Low, Medium and High Levels of Learning

Acquisition Level	Order of Response Execution	Number of Self-Assessment Response Categories			
		Two	Four	Eight	Mean
High	KM	3.306	3.332	3.424	3.354
	MK	3.228	3.316	3.126	3.223
Medium	KM	4.086	4.157	4.418	4.220
	MK	4.302	4.083	4.606	4.330
Low	KM	4.254	4.579	4.805	4.546
	MK	4.630	4.572	5.031	4.744

This finding is consistent with the notion that, during the initial stages of practice at least, the m response is covertly selected first followed by the selection of the level of sureness. As practice proceeds the self-assessment responses may increasingly depend upon some general information memory and/or some direct access process which is not reflected in the response latencies which were measured.

The preceding interpretation assumes that both responses are selected before either is executed. However, one can assume that an m or k response is executed as soon as it has been selected. In this case, and if the m response is selected first, then the MK groups should respond faster than the KM groups because the KM groups involve a step of retrieving a previously selected m response which the MK groups do not require. Thus, the conclusion about the order of internal selection of the m and k responses which one draws from the data critically depends upon which assumption is made. The findings are indefinite with regard to determining the order in which the m and k responses are covertly selected, but it seems reasonable to conclude that the manner in which the m and k responses are processed and/or selected is altered with practice.

Another possibility is that the m and k responses are processed in some parallel fashion--or in some fashion which is more complex than simply

first selecting one response, say the m response, followed by the selection of the other response, say k. Indeed, the proposed model suggests that, first, an m response is tentatively retrieved from a repertoire of m responses, then the correctness of this m response is tested by the individual--which makes available a k response. If the level of k exceeds some criterion level, then the tentatively retrieved m response is selected for execution; otherwise another m response is retrieved, etc.

A separate analysis of the latencies of sure-correct, unsure-correct, sure-wrong, and unsure-wrong responses shows that the effect of O depends upon whether the response was correct or wrong, $F(1, 198) = 4.79$, $p < .05$. The latencies on the low and medium criterion trials and on the one trial before and after each of these criterion trials were all combined for this analysis; and only subjects who made at least one of each of the four kinds of responses were included in the analysis. For correct responses the KM groups completed the two-response sequence approximately 100 msec. faster (4,041 vs. 3,934 msec.) than the MK groups; for the wrong responses the difference was approximately 400 msec. (5,218 vs. 4,811 msec.).

It is also of interest to note that the speed of executing correct responses (but not wrong responses) is affected by the sureness associated with their execution, $F(1, 198) = 51.25$, $p < .01$. The execution of sure-correct responses was accomplished approximately 1 second faster, on the average, than was the execution of unsure-correct responses (3,458 vs. 4,518 msec.), tD(11, 4) = 4.57, $p < .01$. The mean time required to execute wrong responses was approximately 5 seconds (5,015 msec.) and was not significantly affected by the sureness.

It is surprising that when a person indicates that he is "Sure" of the correctness of a response, the wrong responses require approximately 1.5 seconds longer for their execution than do correct responses. This suggests that different processes or different factors are involved in determining the correctness of a response than are involved in determining the overt self-assessment responses. One possibility is that there is a discrepancy between the overt K and the covert k self-assessments. For example, the K responses may be influenced relatively more than the k responses by how rapidly the person thinks he is expected to learn the material.

The Accuracy of the Self-Assessment Responses. One indication of the accuracy of the SA responses is the extent of agreement between the percentage of responses which are correct and the percentage of responses about which the learner is "Sure" of their correctness.

Figure 6 shows the relation between the percentage "Sure" and percentage correct responses for the two-, four-, and eight-category treatments plotted separately for the two orders (MK and KM) in which the M answers and K self-assessment responses were executed. The plotted points were calculated as follows: (1) the number of trials required by each subject to attain the 100% acquisition criterion was divided by 10, (2) the percentage "Sure" and percentage correct responses were determined at each of these 1/10th acquisition points for each subject, using linear interpolation if the points fell between trials, (3) for each subject the correlation, y-intercept, and slope value of the relationship between percentage "Sure"

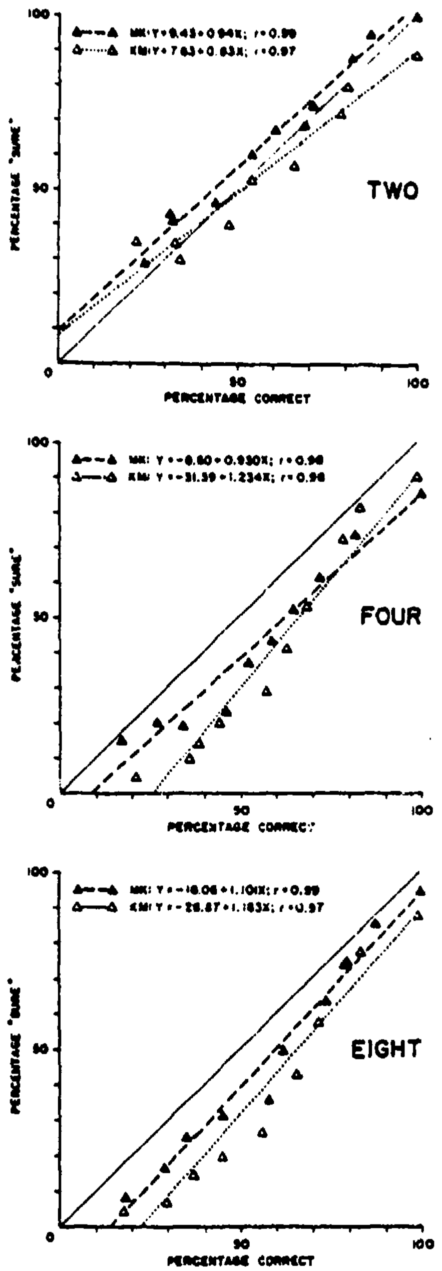

Figure 6. Mean percentage of responses about which the learners were "Sure" of the correctness as a function of the mean percentage of responses which were actually correct when two, four or eight self-assessment response categories were employed. Separate functions are plotted for the groups which executed the answer first (MK) and for the groups which executed the self-assessment first (KM). Each point is the mean of twenty subjects.

and percentage correct was calculated, and (4) the mean values for the 20 subjects in each treatment were calculated.

In a scatter graph of this relationship an infallible self-assessor would be represented by a line from the lower left (0,0) point extending diagonally to the upper right (100, 100) point; the linear correlation (r) between the percentage "Sure" and the percentage correct would be 1.0; the regression line would have a y-intercept value of zero and a slope value of 1.0. This is much like the "perfectly calibrated assessor" described by Lichtenstein and Fischhoff (1976). In Figure 6 values above the diagonal line represent "optimism"--or at least a sureness which exceeds the demonstrated correctness. However, it should be noted that it does not necessarily follow that if all of the points fall on the diagonal then the self-assessments are completely accurate. This is so because the percentage of "Sure" responses may be equal to the percentage of correct responses, but the specific populations of items may be different.

The correlation, y-intercept, and slope values of the obtained relationship between the mean values are shown in each figure. The mean r values range from 0.97 to 0.99, the y-intercept values from -31.6 to +9.4, and the slope values from 0.83 to 1.23. An analysis showed that the slope of the relationship between the percentage "Sure" and the percentage correct was significantly affected by the interaction between the order (O) in which the responses were executed and the number (A) of self-assessment categories, $F(2, 96) = 3.35$, $p < .05$. The K4M treatment has a higher slope than the K2M condition, $tD(6, 96) = 0.377$, $p < .01$, and the MK4 treatment, $tD(6, 96) = 0.339$, $p < .05$. There were no statistically significant differences among the MK treatments.

The analysis of the y-intercept also showed the O x A interaction to be significant, $F(2, 96) = 3.93$, $p < .05$. There was no difference between the MK and KM treatments when only two self-assessment response categories were used. However, the KM treatments showed a lower y-intercept than the MK treatments when either four or eight response categories were employed. The correlation value was not significantly affected by either O or A.

Another indication of the accuracy of the self-assessment responses can be obtained by observing the proportion of responses which are correct if the person states that he is sure (or unsure) of the correctness. These conditional proportions are shown in Table 6 for the different experimental treatments and two levels of acquisition.

Overall, the probability of a correct response was higher if the learner indicated that he was sure (0.72) than if he indicated that he was unsure (0.42) of its correctness, $F(1, 324) = 211$, $p < .01$. This predictive accuracy of self-assessments was greater if the answer was given before the self-assessment (0.75 if sure vs. 0.39 if unsure) than after the self-assessment (0.68 vs. 0.44), $F(1, 324) = 8.46$, $p < .01$; and the accuracy of the self-assessments improves with practice, $F(1, 324) = 4.03$, $p < .05$.

A separate analysis of the responses about which the learner was sure of their correctness showed that the proportion correct was affected by the number of self-assessment categories (A), $F(2, 108) = 3.57$, $p < .05$ and the

interaction of A and L, $F(2, 108) = 3.71$, $p < .05$. The relative disadvantage of using a small number of self-assessment categories decreases with practice. A similar analysis of the responses about which the learner was unsure revealed no significant effect of A.

Table 6

Conditional Proportion of Responses Which Are Correct If a Sure of If an Unsure Self-Assessment Response Is Made

Level of Acquisition	Order of Response Execution	P(Correct | Unsure) Number of Response Categories Two	Four	Eight	P(Correct | Sure) Number of Response Categories Two	Four	Eight
Medium	MK	.434	.537	.402	.839	.748	.903
	KM	.424	.572	.520	.754	.823	.870
Low	MK	.258	.347	.354	.623	.701	.677
	KM	.358	.404	.376	.388	.637	.622

A measure of how accurately the subjects were able to assess the correctness of the m responses before they had executed them (or had executed them but had not yet been informed about their correctness) may be obtained also by employing a signal detection analysis. The general idea is that an internal weak signal may be present within the person which indicates that the tentatively selected m response will produce the desired consequences (or being correct). If the selected m response will not produce the desired consequences then the signal is absent. This internal signal of knowing is simply added (in the same dimension along which the noise randomly varies) to the background of internal noise which accompanies the state of not knowing the correct answer.

Assume that the execution of correct and wrong M answers provide reasonable indications that the individual knows or does not know the correct answer, respectively. And assume that the self-assessment response of "Sure" and unsure provide reasonable estimates of the person's decision that the signal was present or absent, respectively. Then the accuracy of the self-assessment responses may be estimated by calculating the hit rates and false alarm rates based upon the conditional probabilities of p(Sure | Correct) and p(Sure | Wrong), respectively. These conditional probability values for each of the six treatments are presented in Table 7 for the low and medium levels of acquisition. If either the M or K response was not executed within the 8-second time limit which was imposed on the subjects during the experiment then the item was not included in the calculation.

A measure, d', of the sensitivity with which the subjects were able to distinguish between knowing and now knowing the correct answer is also presented in Table 7; these d' values were calculated based upon the mean hit

rates and false alarm rates of the 20 subjects who performed under a particular treatment. Within the proposed model of self-assessment, d' might represent the difference between the mean of a noisy criterion of "sureness" and the mean of a distribution which represents the signal plus the noise.

Table 7

Hit Rate (H) Estimated by P(Sure | Correct), False Alarm Rate (FA) Estimated by P(Sure | Wrong) and the d' Value for Each of the Six Treatments Involving Self-Assessment at Two Levels of Learning

Acquisition Level	Order of Response Execution		Number of Self-Assessment Response Categories: Two	Four	Eight	Mean
Medium	MK	H	.83	.58	.69	.70
		FA	.52	.27	.31	.37
		d'	.90	.81	1.00	.90
	KM	H	.70	.52	.58	.61
		FA	.55	.28	.37	.40
		d'	.39	.63	.53	.52
Low	MK	H	.72	.46	.41	.52
		FA	.42	.14	.16	.24
		d'	.78	.90	.76	.81
	KM	H	.46	.31	.33	.37
		FA	.41	.10	.13	.21
		d'	.13	.78	.69	.53

Inspection of the d' values in Table 7 suggests that subjects can determine more sensitively whether or not they know a correct answer if the M answer is executed first (d' = 0.76 to 1.00) rather than if the K self-assessment is executed first (d' = 0.13 to 0.78). Furthermore, it appears that the greater sensitivity can be attributed to a higher hit rate for the MK groups (0.41 to 0.83) compared to the KM groups (0.31 to 0.70). There seems to be little difference in the false alarm rates for the MK groups (0.14 to 0.52) and for the KM groups (0.10 to 0.55).

A statistical analysis of the hit rates and false alarm rates essentially supports the above observations. An analysis of variance of the hit rates revealed the main effect of $\underline{O}$ to be significant, $\underline{F}(1, 108) = 8.08$, $\underline{p} < .01$. A similar analysis of the false alarm rates produced no statistically significant effect of $\underline{O}$, $\underline{F} < 1.0$.

These findings are consistent with the notion that the execution of an M response provides additional cues which permit the individual to refine or alter a covert k response before executing it. However, the execution of an M response seems to have a differential effect on the self-assessment depending upon whether a correct or wrong M response has been selected and executed. If a correct response has been selected then its execution confirms its correctness and permits the individual to be more sure; this is shown by the higher hit rate for those subjects under the MK treatments. But the execution of a wrong M response apparently provides little or no information to the individual (prior to the receipt of knowledge of results about the real-world consequences) which aids in modifying the self-assessment; the false alarm rate is unaffected by the order in which the M and K responses are executed.

The use of signal detection theory in this analysis assumes, as stated earlier, that a wrong response by the learner indicates validly that he does "not know" it, i.e., that the cues observed by the learner upon which the self-assessment responses are based were produced by "noise alone." However, the learner could also make a wrong response by failing to retrieve or failing to execute properly a response--even though the correct response may be stored in memory. Thus, these estimates may be biased. To the extent that the self-assessment is based only upon whether the answer is stored, there may be an overestimate of the false alarm rate and an underestimate of the hit rate.

GENERAL DISCUSSION

This paper is concerned with the ability of people to assess the correctness of responses which they are anticipating making or have just made (and not yet received knowledge of results), some ways in which the performance of such self-assessments may interact with learning, and with the processes which underlie self-assessment performance. It was found that the performance of a self-assessment task during learning may expedite the rate at which the correct responses in a paired-associates task are acquired, relative to groups of learners who were not required to perform a self-assessment task.

An especial benefit to learning seems to occur if the response to-be-learned is executed prior to the execution of the self-assessment response and if a sufficiently precise self-assessment response is required. In this study the subjects who used either four or eight response categories to make their self-assessment showed more rapid acquisition than the subjects who simply learned the material with no self-assessment.

Also, statistical comparisons restricted to those treatments which involved the self-assessment task showed clearly that the subjects who executed their answer before indicating their self-assessment learned the material in fewer trials than subjects who indicated their self-assessment first. There are several possible interpretations of this finding.

One interpretation involves the finding that the subjects who executed their answer first showed a greater ability to identify a correct response. For example, the probability of saying "Sure" when a correct response was

made was higher under the MK treatments (0.52 to 0.70) than under the KM treatments (0.37 to 0.61). There was little difference between the MK and KM treatments in terms of saying "Sure" when a wrong response was made.

The proportion of the total responses which were both sure and correct at the low acquisition level was higher for the MK treatments (0.22 to 0.42) than for the KM treatments (0.14 to 0.29). Also, the y-intercept values for the four- and eight-category groups were closer to zero under the MK treatments than the KM treatments, which suggests more accurate self-assessments by the learners in MK treatments early in practice.

This greater ability to identify a correct response under the MK treatments may provide, in effect, more and/or quicker feedback information for the subjects under those treatments--and perhaps at an earlier stage of practice.

However, the notion that the subjects in the MK treatments receive more and/or quicker feedback because of their relatively greater accuracy of their self-assessments is arguable. This is so because the subjects under the MK treatments took approximately 200 msec. longer (at the low level of acquisition) to complete the M-K response sequence than did the subjects in the KM treatments require to complete their K-M response sequence. Thus, it is possible that the subjects under the KM treatments covertly altered their self-assessments, within 200 msec., after the execution of the M response, e.g., in the fashion <u>K</u>-<u>M</u>-<u>k</u>. Our data do not resolve this argument.

Another possible interpretation of the more rapid learning under the MK treatments is that the subjects in the MK treatments are relatively greater beneficiaries of reinforcements which are associated with making accurate self-assessment. That is, making an accurate self-assessment in addition to making a correct M response may be more rewarding than making a correct response alone. In support of this reinforcement hypothesis it was found that those responses (correct and wrong combined) which were accurately assessed were repeated more often (53%) than responses which were inaccurately assessed (31%). This tendency to repeat accurately assessed responses was especially apparent for correct responses; 85% of the sure-correct responses were repeated, while only 49% of the unsure-correct responses were repeated.

As stated above, the subjects under the MK treatments showed more accurate self-assessments in a number of ways, e.g., at the low level of acquisition the MK treatments had a higher hit rate (.52 vs. .37), a higher mean d' value (.81 vs. .53), and a higher proportion of sure-correct responses. Thus, according to the reinforcement interpretation, these MK subjects received a greater number of rewards.

> The man who thinks himself worthy of less than he is really worthy of is unduly <u>humble</u>, whether his deserts be great or moderate, or his deserts be small but his claims yet smaller.

. . . he who is worthy of little and thinks himself worthy of little is <u>temperate</u>.

Aristotle, 4th century B.C.
(translator W. D. Ross in Auden, 1970)

REFERENCES

Adams, J. A. A closed-loop theory of motor learning. Journal of Motor Learning, 1971, 3, No. 2, 111-149.

Attneave, F. How do you know? American Psychologist, 1974, 29, No. 7, 493-499.

Auden, W. H. (ed.). The Portable Greek Reader. Clinton, Mass.: The Colonial Press, 1970.

Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 1977, 84, No. 2, 191-215.

Bernbach, H. A. Decision processes in memory. Psychological Review, 1967, 74, 462-480.

Bobrow, D. G. Dimensions of representation. In D. G. Bobrow and A. Collins. Representation and Understanding: Studies in Cognitive Science. New York: Academic Press, Inc., 1975.

Bradley, J. V. Distribution-Free Statistical Tests. New Jersey: Prentice-Hall, Inc., 1968.

Broadbent, D. E. In Defence of Empirical Psychology. Great Britain: Methuen & Co., Ltd., 1973.

Broadbent, D. E., & Broadbent, M. H. P. Effects of recognition on subsequent recall: Comments on "Determinants of recognition and recall: Accessibility and generation" by Rabinowitz, Mandler, & Patterson. Journal of Experimental Psychology: General, 1977, 106, 330-335.

Buchwald, A. M. Effects of "right" and "wrong" on subsequent behavior: A new interpretation. Psychological Review, 1969, 76, No. 2, 132-143.

Carpenter, C. R., Eggleton, R. C., John, F. T., & Cannon, J. R., Jr. The classroom communicator. Technical Report SDC-269-7-14, Human Engineering Branch, Special Devices Center, Fort Washington, L. I., N.Y., 1950.

Deese, J. Behavior and fact. American Psychologist, 1969, 24, 515-522.

d'Ydewalle, G., & Eeleen, P. Repetition and recall of "right" and "wrong" responses in incidental and intentional learning. Journal of Experimental Psychology: Human Learning and Memory, 1975, 104, No. 4, 429-441.

Green, D. M., & Swets, J. A. Signal Detection Theory and Psychophysics. New York: John Wiley and Sons, Inc., 1966.

Hart, J. T. Recall, recognition, and the memory-monitoring process. Stanford University Dissertation, 1966.

Jagacinski, R. J., & Miller, R. A. Describing the human operator's internal model of a dynamic system. Human Factors, 1978, 20(4), 425-433.

Kanfer, F. J. The maintenance of behavior by self-generated stimuli and reinforcement. In A. Jacobs and L. B. Sachs. The Psychology of Private Events: Perspectives on Covert Response Systems. New York: Academic Press, 1971.

Kelley, C. R. Manual and Automatic Control: A Theory of Manual Control and Its Application to Manual and to Automatic Systems. New York: John Wiley and Sons, Inc., 1968.

Kelso, J. A. S. Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 529-543.

Kolers, P. A., & Palef, S. R. Knowing not. Memory & Cognition, 1976, 4(5), 553-558.

Levine, M. A Cognitive Theory of Learning: Research on Hypothesis Testing. New York: John Wiley and Sons, 1975.

Levine, J. M., Samet, M. G., & Brahlek, R. E. Information seeking with input pacing and multiple decision opportunities. Human Factors, 1974, 16, 384-394.

Lichtenstein, S. C., & Fischhoff, B. Do those who know more also know more about how much they know? Oregon Research Institute Bulletin, 16, No. 1, Feb. 1976, Eugene, Oregon 97403, PO Box 3196.

Miles, J. Eliminating the guessing factor in the multiple choice test. Educational and Psychological Measurement, 1973, 33, 637-651.

Miller, G. A., Galanter, E., & Pribram, K. H. Plans and the Structure of Behavior. New York: Holt, Rinehart and Winston, Inc., 1960.

Murdock, B. B., Jr., & Dufty, P. O. Strength theory and recognition memory. Journal of Experimental Psychology, 1972, 94, No. 3, 284-290.

Nuttin, J., & Greenwald, A. G. Reward and Punishment in Human Learning. London: Academic Press, Inc., Ltd., 1968.

Rabinowitz, J. C., Mandler, G., & Patterson, K. E. Determinants of recognition and recall: Accessibility and generation. Journal of Experimental Psychology: General, 1977, 106, 302-329.

Roy, E. A., & Marteniuk, R. G. Mechanisms of control in motor performance: Closed-loop versus motor programming control. Journal of Experimental Psychology, 1974, 103, 985-991.

Sampson, J. R., & Chen, I. N. Computer simulation of multiple strategies in human binary prediction. Psychological Reports, 1971, 29, 1207-1221.

Shuford, E.H., & Massengill, H.E. The relative effectiveness of five instructional strategies. Interim Technical Report, Contract No. AF 49(638) - 1744, ARPA Order No. 833, 1967.

Siegel, S., (Siegel, E., & Andrews, J. M). Choice, Strategy, and Utility. New York: McGraw-Hill Book Company, 1964.

Spear, N. E. The Processing of Memories: Forgetting and Retention. New York: John Wiley and Sons, 1978.

Tolman, E. C. Principles of purposive behaviorism. In S. Koch (Ed). Psychology: A Study of a Science Vol. 2, General Systematic Formulations, Learning, and Spatial Processes. New York: McGraw Hill Book Company, Inc., 1959.

Tulving, E. Cue-dependent forgetting. American Scientist, 1974, 62 74-82.

Wade, T. C. Relative effects on performance and motivation of self-monitoring correct and incorrect responses. Journal of Experimental Psychology, 1974, 103, No. 2, 245-248.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Co., 1971.

Woodfield, A. Teleology. Cambridge (England): Cambridge University Press, 1976.

Woodworth, R. ~~S.,~~ & Schlosberg, H. Experimental Psychology. New York: Holt, Rinehart and Winston, 1954.

DISTRIBUTION

2 HQDA RESEARCH AND STUDIES OFC
22 ARI LIAISON OFFICE
1 INTER-UNIV SEMINAR ON ARMED FORCES + SOC
1 OASA (RDA) DEPUTY FOR SCIENCE AND TECHNOLOGY
1 OFC OF NAVAL RESEARCH /
1 AFHRL/LRT
1 AFHRL/LRLG
1 AIR FORCE HUMAN RESOURCES LAB ATTN: AFHRL/TSR
1 NAVY PERSONNEL R AND D CENTER DIRECTOR OF PROGRAMS
1 NAVY PERSONNEL R AND D CENTER /
2 OFC OF NAVAL RESEARCH PERSONNEL AND TRAINING RESEARCH PROGRAMS
1 NAVAL PERSONNEL R + D CENTER /
1 OFC OF NAVAL RESEARCH PROJECT OFFICER, ENVIRONMENTAL PHYSIOLOGY
1 NAVAL AEROSPACE MEDICAL RSCH LAB AIRBORNE RANGER RESEARCH
1 NAVAL AEROSPACE MEDICAL RSCH LAB AEROSPACE PSYCHOLOGY DEPARTMENT
1 HEADQUARTERS, COAST GUARD CHIEF, PSYCHOLOGICAL RSCH BR
1 USA RESEARCH AND TECHNOLOGY LAB /
1 NAFEC HUMAN ENGINEERING BRANCH
1 HQ WRAIR DIV OF NEUROPSYCHIATRY
1 USA RSCH DEVEL + STANDARDIZA GP, U.K.
1 USA RESEARCH AND DEVELOPMENT LABS CHIEF, BEHAV SCIENCES DIV, FOOD SCI LAB
1 UNIFORMED SERVICES UNIT OF THE HEALTH SCI DEPARTMENT OF PSYCHIATRY
1 CENTER FOR NAVAL ANALYSIS
1 NAVAL HEALTH RSCH CEN LIBRARY
1 NAVAL PERSONNEL R AND D CEN LIBRARY ATTN: CODE P106
1 AIR FORCE HUMAN RESOURCES LAB ATTN: AFHRL/OTS
1 USA ACADEMY OF HEALTH SCIENCES STIMSON LIBRARY (DOCUMENTS)
1 ERIC PROCESSING AND REFERENCE FAC ACQUISITIONS LIBRARIAN
1 USMA DEPT OF BEHAVIORAL SCI AND LEADERSHIP
1 OLD DOMINION UNIVERSITY PERFORMANCE ASSESSMENT LABORATORY
1 NMRDC PROGRAM MANAGER FOR HUMAN PERFORMANCE
1 U OF TEXAS CEN FOR COMMUNICATION RSCH
1 INSTITUTE FOR DEFENSE ANALYSES
1 AFHRL TECHNOLOGY OFC (H)
1 PURDUE UNIV DEPT OF PSYCHOLOGICAL SCIENCES
1 USAFA DEPT OF BEH SCI + LEADERSHIP
1 US MILITARY ACADEMY DEPT. OF HISTORY, BLDG 601
1 USA INTELLIGENCE CEN AND SCH ATTN: SCHOOL LIBRARY
1 US MILITARY ACADEMY DIRECTOR OF INSTITUTIONAL RSCH
1 NAVAL POSTGRADUATE SCH ATTN: DUDLEY KNOX LIBRARY (CODE 1424)
1 HQ ATC/XPTD TRAINING SYSTEMS DEVELOPMENT
1 USA FIELD ARTILLERY SCHOOL MORRIS SWETT LIBRARY
1 USA INSTITUTE OF ADMINISTRATION ACADEMIC LIBRARY
1 USA WAR COLLEGE ATTN: LIBRARY
1 USA ENGINEER SCHOOL LIBRARY AND LEARNING RESOURCES CENTER
1 USA ARMOR SCHOOL (USARMS) ATTN: LIBRARY
1 US COAST GUARD ACADEMY LIBRARY
1 ORGANIZATIONAL EFFECTIVENESS CEN + SCH ATTN: LIBRARIAN
2 BRITISH EMBASSY BRITISH DEFENCE STAFF
1 COLS (W) LIBRARY
2 CANADIAN FORCES PERSONNEL APPL RSCH UNIT
1 ARMY PERSONNEL RESEARCH ESTABLISHMENT
6 LIBRARY OF CONGRESS EXCHANGE AND GIFT DIV
1 DEFENSE TECHNICAL INFORMATION CEN ATTN: DTIC-DDA-2
140 LIBRARY OF CONGRESS UNIT DOCUMENTS EXPEDITING PROJECT
1 US GOVERNMENT PRINTING OFC LIBRARY, PUBLIC DOCUMENTS DEPARTMENT
1 US GOVERNMENT PRINTING OFC LIBRARY AND STATUTORY, LIB DIV (SLL)
1 THE ARMY LIBRARY ATTN: ARMY STUDIES SEC